BELLCOMM, INC.
1100 SEVENTEENTH STREET, N.W.   WASHINGTON, D.C.   20036

COVER SHEET FOR TECHNICAL MEMORANDUM

TITLE- Statistical Characteristics of
Solar Proton Events

TM- 67-1033-2

DATE- July 1, 1967

FILING CASE NO(S)- 340

AUTHOR(S)- P. Gunther

FILING SUBJECT(S)-
(ASSIGNED BY AUTHOR(S)- Radiation
Solar Events

ABSTRACT

        The statistical quality of the data on solar proton events
is extremely poor, so that consideration only of general statistical
characteristics appears warranted. The principal conclusions are:

1. Solar events tend to cluster within two - six days of each
    other. This will affect the maximum short-time radiation
    exposure.

2. The data suggests, although not conclusively, that there
    is a maximum "cutoff" flux for single events.

3. The annual flux follows, in general outline, the well-
    known variation of sunspots within the eleven-year solar
    cycle. Hence, expected cumulative flux will depend on the
    dates encompassed by any particular mission. Detailed
    statistical analysis of historical and current sunspot
    data may lead to significantly improved predictions for
    the present cycle.

        Perhaps the main result of the analysis is that, in order
to obtain valid statistical estimates, it is essential that adequate
statistical controls be incorporated in any information gathering
bearing on solar radiation. It is recommended, therefore, that the
statistical point of view be represented during the formative stages
of program planning in this area.

        An appendix presents a tutorial exposition of confidence
distributions, with applications to various types of population
distributions. Using a simple model, confidence estimates of the
maximum cutoff flux for a single event are derived

BA-145A (3-67)

DISTRIBUTION

COMPLETE MEMORANDUM TO	COVER SHEET ONLY TO

CORRESPONDENCE FILES:

OFFICIAL FILE COPY
plus one white copy for each
additional case referenced

TECHNICAL LIBRARY (4)

COMPLETE MEMORANDUM TO	COVER SHEET ONLY TO
F. B. Benjamin - NASA/MM	F. G. Allen
P. E. Culbertson - NASA/MLA	G. R. Andersen
F. P. Dixon - NASA/MTY	G. M. Anderson
E. W. Hall - NASA/MTS	A. P. Boysen, Jr.
T. A. Keegan - NASA/MA-2	A. N. de Gaston
J. W. Keller - NASA/RV-1	J. P. Downs
D. R. Lord - NASA/MTD	P. L. Havenstein
W. C. Mathews - NASA/ML	D. R. Hagner
M. J. Raffensperger - NASA/MTE	H. A. Helm
A. Reetz, Jr. - NASA/RV-1	W. C. Hittinger
L. Reiffel - NASA/MA-6	B. T. Howard
A. D. Schnyer - NASA/MTV	D. B. James
J. H. Turnock - NASA/MA-4	A. N. Kontaratos
	K. E. Martersteck
J. C. Lill - MSC/EF2	J. Z. Menard
J. L. Modisette - MSC/EF2	I. D. Nehama
D. Robbins - MSC/TG43	G. T. Orrok
	C. A. Pearse
	T. L. Powers
	I. M. Ross
	J. M. Tschirgi
	T. C. Tweedie
	R. L. Wagner
	Department 1023

BA-146 (8-64)

BELLCOMM, INC.
1100 SEVENTEENTH STREET, N.W.   WASHINGTON, D.C.   20036

COVER SHEET FOR TECHNICAL MEMORANDUM

TITLE- Statistical Characteristics of
Solar Proton Events

TM- 67-1033-2

DATE- July 1, 1967

FILING CASE NO(S)- 340

AUTHOR(S)- P. Gunther

FILING SUBJECT(S)-
(ASSIGNED BY AUTHOR(S)- Radiation
Solar Events

ABSTRACT

The statistical quality of the data on solar proton events is extremely poor, so that consideration only of general statistical characteristics appears warranted. The principal conclusions are:

1. Solar events tend to cluster within two - six days of each other. This will affect the maximum short-time radiation exposure.

2. The data suggests, although not conclusively, that there is a maximum "cutoff" flux for single events.

3. The annual flux follows, in general outline, the well-known variation of sunspots within the eleven-year solar cycle. Hence, expected cumulative flux will depend on the dates encompassed by any particular mission. Detailed statistical analysis of historical and current sunspot data may lead to significantly improved predictions for the present cycle.

Perhaps the main result of the analysis is that, in order to obtain valid statistical estimates, it is essential that adequate statistical controls be incorporated in any information gathering bearing on solar radiation. It is recommended, therefore, that the statistical point of view be represented during the formative stages of program planning in this area.

An appendix presents a tutorial exposition of confidence distributions, with applications to various types of population distributions. Using a simple model, confidence estimates of the maximum cutoff flux for a single event are derived
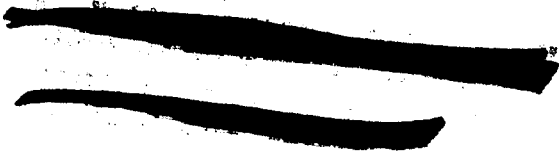
BELLCOMM, INC.

TECHNICAL MEMORANDUM

I. INTRODUCTION

Radiation from solar proton events is a potentially serious hazard to long duration manned space flights. Permissable radiation doses (see reference 1) are ordinarily specified both in terms of maximum accumulated dose for the entire mission, as well as maximum individual dose in one day, i.e., a single event. For given spacecraft shielding configuration and mission duration, it becomes necessary then to estimate the radiation levels that would not be exceeded with high probability, e.g., 99%; moreover, the confidence in such an estimate should also be high, e.g., 95%. This memorandum considers only the question of estimating the accumulated and individual fluxes arising from solar proton events; converting these into resulting doses for different shielding configurations is not considered.

Unfortunately the data from the last (19) solar cycle exhibits large fluctuations and numerous statistical discrepancies, so that a detailed quantitative statistical analysis is unlikely to be meaningful. Emphasis here is placed primarily on qualitative statistical characteristics, particularly those features likely to affect the choice of model. In fact, a principal objective is to point out explicit shortcomings of the data and the consequent statistical difficulties that arise. It is hoped thereby that the statistical requirements for data collection during the present solar cycle can be better understood and the various problems likely to be encountered in a quantitative statistical analysis can be anticipated.

Section 2 discusses the limitations of the data. Section 3 summarizes the salient statistical characteristics and section 4 presents a summary and conclusions. The appendix discusses quantitative statistical procedures and presents illustrative calculations for a very simple model.

II. DISCUSSION OF QUALITY OF THE DATA

2.1 Resume of Data Sources

This section briefly summarizes the types of measurements used in estimating flux. A more detailed discussion is given in reference 2.

When a solar event occurs, the impacting particles cause an increase in ionization of the atmosphere. This in turn increases the absorption of any radio waves that impact the ionosphere. Estimates of flux from ground based observations are derived from theories which relate the incident flux to the change in radio wave intensity. The two principal ground based networks for observing this phenomenon employ (a) the riometer, or (b) the forward scatter method.

The riometer is a passive system. Upon the occurrence of a solar event, it measures the amount of attenuation in the normal background galactic noise, the lower energy limit being about 5 Mev. It is estimated (reference 2) that there is about a 20% experimental uncertainty in calculating the amount of absorption, and an additional factor of 2 uncertainty exists in interpreting the data to arrive at an estimated flux. The background noise is such that small events would not ordinarily be detected. The riometer data is ordinarily supplemented by neutron monitors which furnish information for energies greater then 1 Bev. In addition, at the start of events, balloons have been deployed at various latitudes to gather additional data. Spectral characteristics and time dependence of the event can generally be inferred from the combination of data.[*]

In the forward scatter method, radio waves are transmitted to the ionosphere and reflected back to a receiver. When a solar event occurs, both the reflection as well as the absorption of these radio waves are increased. This method appears to have greater uncertainty than the riometer. In addition to day-night variation, the measurements are affected by changes in the earth's magnetic field. The existence and magnitude of a solar event are determined only after extended study of the normal long term background radio signal reception characteristics. One of the principal investigators employing the forward scatter method is Bailey. (Bailey's compilation[4] is

---

[*]Although the time variation is important in considering different types of radiation symptoms, the event is usually summarized by estimating the total flux integrated over the duration of the event.

referred to below as Bailey events.) On the other hand, Webber's compilation[3] (called Webber events) for the most part discounts the forward scatter data.

Satellites, which operate above the atmosphere, offer the best means of observing events since they detect the primary particles directly. The satellite data is discussed in the following section 2.2.

The flux data for the 54 observed solar events of the nineteenth solar cycle is presented in Table 1 (from reference 5). A summary tabulation is given in Table 2.

2.2 Satellites and Occurrences of Small Events

The Explorer VII satellite was operational from October 13, 1959 to February 17, 1961. During this period there was a large increase in the number of recorded small events. This is shown in the summary table below for energies > 30 Mev.

ANNUAL NUMBER OF SMALL EVENTS

	(Webber events) Flux between 10^6 and 10^7	(Webber events) Flux less than 10^6	(Bailey events) All
1956	0	0	2
1957	3	0	5
1958	2	0	1
1959	1	0	1
1960	8	3	1
1961	3	0	1

Actually reference 6 lists 21 events observed by Explorer VII from October 13, 1959 to February 17, 1961. Fourteen of these are included in Webber's list[3], while seven events (mostly very small or occurring shortly after other events) appear in neither Webber's nor Bailey's list. One Webber event (September 26, 1960) and one Bailey event (March 29, 1960) were not reported by Explorer VII.

It appears likely that if a satellite had been operational prior to October 1959, many more small events would probably have been detected. Webber[3] estimates 100 events as a likely total for the nineteenth solar cycle--this is almost twice the number (54) listed in Table 1.

2.3 Bailey's Events

Some difference of opinion exists regarding the interpretation of events reported by Bailey. Such events have been included in the analysis of reference 5 and excluded in reference 2. Of the 54 events listed in Table 1, 34 of these are included by both Webber and Bailey; nine are included by Webber but not by Bailey; and 11 are included by Bailey but not by Webber.* For these latter 11 events, Bailey's flux estimates appear to be surprisingly large by comparison; all are greater than 10^7, the largest (March 25, 1958) being 7.8 x 10^8.

The question arises of how to treat the Bailey events statistically. To eliminate them completely might well bias further the number of actual events, already undoubtedly underestimated. On the other hand, in view of the obvious discrepancies in flux magnitudes between Webber and Bailey events, inclusion of the latter would add appreciably to the already considerable amount of "noise". A reasonable compromise--which will be adopted in subsequent analysis--is to include the Bailey events but to reduce the fluxes. We have more or less arbitrarily used a reduction factor of 100. In effect, this places all Bailey events in the small event category.

In summary, there appears to be large uncertainties in the data, including a probable underestimate in the number of small events (except possibly for 1960). There are also discrepancies between Webber and Bailey events both in identification as well as estimation of flux magnitude.

III. STATISTICAL CHARACTERISTICS OF THE DATA

This section presents a qualitative statistical analysis of the data in Table 1. Section 3.1 discusses the clustering effect, i.e., the tendency for many events to occur within a few days of one another. Section 3.2 compares the flux data corresponding to energies greater than 30 Mev with energies greater than 100 Mev. Section 3.3 discusses the annual variation in flux, especially in connection with the 11-year cycle describing variation in sunspot activity. Section 3.4 combines the individual

---

*Webber does include, in a later portion of his report (Table 8 in reference 4), four of Bailey's events--March 10, 1956; November 13, 1956; April 3, 1957; and June 22, 1957; however, no estimates of flux are given.

events for all years and discusses the appropriateness of a
variety of alternative statistical distributions that might be
used to describe the observed fluxes for individual events
and/or clusters.

It is useful initially to note that most of the
statistical inferences and conclusions are influenced primarily
by the four largest events (or clusters) that occurred. In
approximate order of importance these are:

FLUXES FOR LARGEST EVENTS

Date	N(>30 Mev) Protons/cm^2	N(>100 Mev) Protons/cm^2
7-10-59	1.0 x 10^9	1.4 x 10^8
7-14-59	1.3 x 10^9	1.0 x 10^8
7-16-59	9.1 x 10^8	1.3 x 10^8
11-12-60	1.3 x 10^9	2.5 x 10^8
11-15-60	7.2 x 10^8	1.2 x 10^8
11-20-60	*	*
2-23-56	1.0 x 10^9	3.5 x 10^8
5-10-59	9.6 x 10^8	8.5 x 10^7

3.1 Clustering

An important characteristic of the solar events is the
frequent occurrence of clusters. In Table 1 braces are used to
identify those cases for which the time interval between succes-
sive events was seven days or less. Nine such clusters are shown,
comprising 28 of the total of 54 events. The following table, which
summarizes the distribution of waiting times between events,
shows that a disproportionately large number of events occurred

---

*Medium size event

with time intervals less than seven days, much more so than would reasonably occur by chance alone.*

DISTRIBUTION OF WAITING TIMES BETWEEN EVENTS

Time Between Events (Days)	Number of Events	
1	2	
2	6	
3	3	
4	3	
5	2	
6	2	
7	1	
8 - 10	0	(19)
11 - 20	8	
21 - 30	6	(14)
31 - 50	6	
51 - 70	4	(10)
71 - 90	5	
91	1	
144	1	
146	1	
172	1	
234	1	(10)

*With 53 events (excluding the first) occurring in 2044 days, the average waiting time is 38.6 days. If events were random; i.e., exponentially distributed, the expected number of events with intervals of four days or less would be 53 [1 - exp (-4/38.6)] = 5.06 compared with 14 events actually observed. Similarly for seven days or less the expected number is 8.8 and the observed number 19. This calculation does not take into account the 28-day period of solar rotation. Since about 75% of the observed protons came from a small longitudinal region of the sun,[16] one would expect somewhat fewer events separated by intervals of 7-21 days.

Thus the statistical evidence suggests some type of triggering or chain reaction effect. If so, one might expect events within a cluster to arise from the same general solar geographic region. Taking into account the 13°/day solar rotation (28-day period), Table 1 indicates (by +) those events within a bracket which can be geographically correlated, and (by -) those which are not.* Estimates of geographic location for the Bailey events were not available.

Of the nine clusters, six involve more than one Webber event. Four of these are completely correlated geographically, including the two very large clusters of events occurring in July 1959 and November 1960. The first of the three August 1958 events does not appear to be correlated. Also, for the cluster of five events in 1960 between April 28 and May 13, only the first and third appear to be geographically correlated.

Of course, by chance alone, one would expect successive events occasionally to occur within seven days. For purposes of subsequent statistical analysis we will adopt the rule that successive events not more than seven days apart constitute part of a cluster only if they show geographic correlation. For Bailey events it will be assumed that all bracketed events form part of a cluster. This encompasses the two events in April 1957, the March 1958 event, and also the two 1957 events on August 31 and September 2.

One should note the statistical correlation between magnitudes of flux for events within a cluster. For the largest cluster of July 1959, the correlation, for both 30 Mev and 100 Mev fluxes, is almost 100 percent. However, the November 1960 and July 1961 clusters had considerably more variation in flux magnitudes.

The clustering effect has important implications both for the statistical analysis and for interpretation of single event dosage. Regarding the latter, solar events occurring within one week lead to radiation symptoms intermediate to single day and long term exposure (see reference 1). From the statistical standpoint one should ascertain whether the individual event or the individual cluster provides a more satisfactory overall description of solar event activity. If the cluster is adopted as the basic unit, the further question arises whether one should use the average, the maximum, or the total flux of all the events within a cluster. Although each of these measures leads to somewhat different statistical results, only the total is analyzed in section 3.3.

---

*The determination of pluses and minuses was made by R. Hilberg, Bellcomm.

Although not directly related to the question of clustering there appears to be some indication that the longest waiting times occurred at the end of each calendar year. Listed below are the "transition" times between the last event of any year and the first event of the following year.

	Transition Time (days)	Remarks
1956-57	69	The shortest for all five years. However, almost all waiting intervals between March 10, 1956 and June 22, 1957 are about twice the average.
1957-58	91	By far the longest time for all of 1957 through first half of 1958.
1958-59	144	
1959-60	146	
1960-61	234	

The possibility that less solar event activity takes place during winter months is examined further in section 3.3 from the standpoint of total quarterly flux.

3.2   Comparison of 30 Mev and 100 Mev Data

Particles with energies of about 30 Mev will just penetrate portions of the skin of the command module. As more shielding is employed, the energy required for penetration increases. In order to estimate dose one needs to know more than merely the integrated flux for all particles with energies greater than 30 Mev.

Table 1 presents fluxes above 100 Mev for 29 of the total of 44 Webber events. Approximately 11 of the 15 missing values are for very small events (30 Mev flux less than 9×10^6 protons/cm^2), three events had 30 Mev flux between 9×10^6 and 2×10^7, and one (June 13, 1959) had 30 Mev flux of 8.5×10^7. For purpose of statistical analysis, it seems preferable to estimate these missing values than to omit them. As discussed below, the 100 Mev fluxes so assigned correspond to 1/10 the 30 Mev flux.

Figure 1 shows a joint plot of the 30 Mev and 100 Mev data. The line representing a 10:1 flux ratio is shown, and can be seen to be not unreasonable as an average fit. The figure also shows considerable scatter in the points. Measurement errors undoubtedly comprise a large portion of this scatter. However, there is also the inherent variation in the (integrated) energy-flux distribution.* The variation appears to be most pronounced in the case of intermediate size events which seem to divide into two fairly distinct groups. It seems safe to say that this variability in the energy-flux distribution is a major factor in the statistical reliability of dose estimates; it may in fact prove necessary to incorporate the entire energy-flux distribution into the analysis.

3.3 Solar Proton Events, Sunspots and Solar Flares

The pattern of proton event activity parallels roughly that for sunspots and solar flares. Figure 2 shows de Gaston's estimates[15] for the nineteenth solar cycle of smoothed sunspot number and number of flares. Also shown is the log of total annual flux with energy greater than 30 Mev. Taking into account the large uncertainties in the flux data, especially prior to 1960, the correspondence is on the whole quite good. As one might expect, the flux follows the flare curve somewhat better than the sunspot curve. The flux peak in 1956, one year prior to the solar flare peak, can be partially explained by the fact that the flare curve plots number of flares. Actually three times as many proton events occurred in 1957 than in 1956. The February 23, 1956 event was the first consequential event of the nineteenth solar cycle, and at the same time the largest for energies greater than 100 Mev.

The deviation between sunspot activity and proton event activity would be considerably greater if total quarterly flux (see Table 3) were plotted rather than annual flux. However, as noted previously, the second and third quarters of each year may be more active than the first and fourth quarters. The following table summarizes the biquarterly fluxes for each of the six years. The contrast is seen to be greater with respect to number of events than for total flux. Although the second and third quarters had larger fluxes for only three of the six years, the total for all years is 50% larger (due primarily to 1959).

---

*Geophysicists usually refer to the rigidity (momentum per unit charge) distribution. However, this is readily transformed into an energy distribution.

SUMMARY OF BIQUARTERLY ACTIVITY

	Annual	1st and 4th Quarters		2^{d} and 3^{d} Quarters	
	Flux (>30 Mev) x 10^8	No. Events	Flux (>30 Mev) x 10^8	No. Events	Flux (>30 Mev) x 10^8
1956	10.26	3	10.01	1	.25
1957	4.63	3	2.60	9	2.03
1958	7.50	3	2.70	6	4.80
1959	42.55	1	.003	6	42.55
1960	21.35	5	20.65	11	.70
1961	3.54	0	0.00	6	3.54
	89.83	15	35.96	39	53.87

In view of the correlation between solar proton events and sunspots, a natural course for predicting proton event activity for the next solar cycle would be first to predict the sunspot activity. Otherwise, since proton event data is available only for the nineteenth solar cycle, one can only assume that the characteristics of the next cycle will be the same as the last.

The historical data on smoothed sunspots for the past nineteen cycles is shown in Figure 3 (from reference 7). Indicated below are a number of qualitative statistical characteristics of these data that appear, after more or less cursory examination, to be significant.

1. Each cycle is remarkably close to 11 years. Major deviations occurred only in the third, fourth and ninth cycles.

2. There is substantial variation in the magnitudes of the peaks, but there is also substantial correlation between successive years. Thus, the first four cycles are large, the next three small, the next four large, the next five small, and the final three large.

3. For the years when peaks were large, the corresponding activity at solar minimum also tended to be relatively high. Moreover, the time when the minimum occurred

appears to have been delayed somewhat. Of the two cases with sharpest drop in activity from one cycle to the next--cycles 4 to 5 and 11 to 12--the time between peak and trough was about nine years. On the other hand, this same time interval occurred also between the relatively low activity cycles 13 and 14.

4. The rate of increase in activity (i.e., the slope of the curve) at the beginning of the cycle appears to be much sharper than the fall off in the latter portion. (The only exception is apparently the seventh cycle.) Moreover, cycles with high activity generally had the sharpest initial rise.

5. Double peaks about two years apart occurred in several cases,* while in other instances it appears that a double peak "almost" occurred. (One would have to refer back to the unsmoothed data to make an accurate determination.)

In summary, the historical sunspot and solar flare data, combined with data since 1961 bearing on the present (twentieth) solar cycle, contains much information that might be useful in predicting the level of solar event activity in the near future. Undoubtedly, complex statistical analyses would be required to extract this information. One should note that only in recent years have sophisticated statistical tools been developed for the analysis of stochastic processes (time series).

3.4 Probability Distribution of Flux for Single Events and Clusters

In order to determine the maximum flux which would not be exceeded with high probability (say 99%), it is necessary to extrapolate the observed data. The usual procedure is first to determine an appropriate (parametric) type of distribution, then estimate the parameters of the distribution, and finally determine the confidence in the estimates. Particularly in the present problem, it is important to use as much as possible of the available information in order to increase the confidence in the extrapolation (i.e., decrease the size of the confidence interval). Although the small events are most numerous, the largest events contain the most information concerning the nature of the

---

*A. N. de Gaston[15] shows pronounced double peaks in number of solar flares for both the eighteenth and nineteenth solar cycles.

distribution in the tails, in particular, whether there is likely to be an absolute maximum flux for individual events.

The most commonly used statistical approach for determining the appropriate type of distribution (e.g., normal distribution) is to plot the observed statistical empirical distribution on probability paper. (Probability paper is scaled so that if there were no random fluctuations the observations would lie on a straight line.) Implicit in this procedure are the following three assumptions:

    i.   the observations are mutually independent,

    ii. they are identically distributed, and

    iii. the distribution is not truncated.

Actually, none of these conditions are fully met in the present situation. Re (i), the presence of clustering (see 3.1) implies that events within a cluster are not independent. Re (ii), the annual variation in activity in accordance with the solar cycle implies that the individual events are not from identical populations. Re (iii), as discussed in section 2, (Webber) events with 30 Mev flux less than 10^6 protons/cm^2 are ordinarily not detected (for 1960 less than 4×10^5). There is also additional uncertainty regarding the calibration of the Bailey events.

In spite of the questionable validity, it may nonetheless be useful to combine all the individual events and examine the resulting probability distribution of fluxes. It seems preferable to work with the relative probability density distribution rather than the integrated cumulative distribution. Besides eliminating the truncation problem, the density actually constitutes a more sensitive representation since any peculiarities in the distribution are not as likely to be smoothed out. On the other hand, it has the disadvantage of inherently providing less smoothing of the random fluctuations--smoothing is accomplished by using a sufficiently large class interval size when plotting the empirical density.

The following analysis considers separately the individual events and the clusters of events, both for the 30 Mev and the 100 Mev data. The principal assumptions, the rationale for which have been previously discussed, are:

    i.   The 30 Mev flux for Bailey events are reduced by a factor of 100.

    ii. Missing 100 Mev fluxes are estimated to be 1/10 the 30 Mev flux.

iii. Successive events occurring within seven days are
assumed to form part of a cluster, provided also that
they arise from the same solar geographic region.

The data are plotted in figures 4 to 7. The observed
fluxes for individual events or clusters (shown at the bottom of
each figure) have been divided into intervals--selected where the
gaps between observations were largest--which then form the esti-
mate of the empirical density. Also shown are the fits obtained
for different types of distributions--uniform, linear, and normal
(the normal for 30 Mev only).* x^2 values which measure goodness
of fit, together with the associated probabilities, have been
computed for each of the fits.** The results are summarized in
Table 4.

---

*The linear density was fit by eye, and the normal density
using the formulas given in reference 8 for the truncated normal
distribution. For simplicity, the sum and sum of squares, needed
to estimate parameters, were calculated from the grouped data.

**x^2 is calculated from the well-known expression

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where   k = total number of intervals used to determine the
empirical density

O_i = observed number of events (or clusters) in the ith
interval

E_i = expected number of events that would occur in the ith
interval using the best estimates of the parameters
of the distribution

The probability associated with x^2 corresponds to the probability
that random sampling fluctuations alone would lead to the value of
x^2 (or larger) calculated from the observed data, provided that the
distribution model assumed were actually the correct one. The
smaller the probability, the less confidence one would place in the
model. The number of degrees of freedom to use in determining the
x^2 probability equals the number of intervals k less the number
of parameters fitted less one (for normalization). It is useful
to note that the average value of x^2 with n degrees of freedom is
n and the standard deviation is $(2n)^{1/2}$. It should be noted that
the method used to select the intervals amounts to a smoothing
operation which biases downward the x^2 values and increases the
probabilities.

The most striking characteristic of the distribution of individual events for 30 Mev (Figure 4) is the bunching of very large events. This shows up as a sharp increase in the density with a sudden fall off to zero, and strongly suggests the presence of a maximum cutoff flux for individual events. The χ^2 calculations tend to support this. The uniform distribution gives the best fit, solely because of the sharp cutoff. The linear density with cutoff at $10^{10.4}$ fits well except at the tails. In fact, when the last two class intervals are combined--which is equivalent to assuming the seven largest events, all near 10^9, to be spread out to $10^{10.4}$-- the χ^2 probability increases from .22 to .996. This result lends further support to the notion of a sharp cutoff flux. A similar result occurs for the truncated normal distribution, although the fit is not as good as for the linear density.

If one examines clusters instead of individual events (Figure 5), the effect is to spread out the large events so that the sharp cutoff is no longer evident. The linear fit (with cutoff at $10^{10.6}$) is quite good and slightly better than the normal.

The 100 Mev plots (Figures 6 and 7) do not present the same picture as the 30 Mev plots. The data are seen to be quite erratic. The uniform density fits very badly and the linear density is rather poor, due primarily to the large gap between the intermediate and the very large events. In contrast to the 30 Mev plots, for large fluxes the density falls off more gradually for individual events (Figure 6) than for clusters (Figure 7).

In summary, although the 30 Mev data lend considerable support to the notion of a maximum cutoff flux for individual events, this conclusion is not fully corroborated by the 100 Mev data. If the data from the twentieth solar cycle turns out to be of sufficiently good quality, it should be possible to resolve this important point.

Since the phenomenon underlying solar events can plausibly be considered as extremal in nature, some remarks may be appropriate regarding the use of an extreme value distribution for the data. There are actually three different types of asymptotic (largest) extremal distributions, depending upon whether the tail of the original random distribution (from which the extremes are taken): (I) falls off at an exponential rate or faster, (II) falls off at a rate slower than exponential, or (III) has a finite cutoff value.*

---

*References 9 and 10 present a good exposition of the theory of extreme value distributions.

Types I and II are two parameter distributions (neither one a cutoff parameter), and it is guessed that the data would fit about as well as for the normal distribution. The type III distribution has three parameters, one being a cutoff parameter, and might be expected to give better results. Unfortunately, statistical procedures for estimating the parameters of truncated extremal distributions have not been developed. Even for the non-truncated case, only ad hoc non-optimal procedures are available. A preliminary analysis of what is believed to be an optimal estimation procedure is presented in section A 5 of the appendix.

The appendix discusses the general statistical problem of determining confidence limits, using the concept of confidence distribution. The approach is essentially equivalent to, but simpler and more general than, the classical statistical method. The uniform, normal, and type III extreme value distributions are discussed, and numerical calculations of confidence limits are presented, when the events can be assumed to be uniformly distributed.

IV. SUMMARY AND CONCLUSIONS

1. Statistically the data on solar proton events is of poor quality, exhibiting many internal inconsistencies. Consequently, none of the usual empirical probability distributions can be confidently used as a basis for extrapolation to high probabilities.

2. For individual events, the fact that all of the largest events had approximately the same flux (for energies > 30 Mev) suggests that a maximum cutoff flux may well exist. However, this conclusion is neither confirmed or rejected by the more erratic flux data for energies > 100 Mev.

3. There is a pronounced tendency for events to cluster. This effect was most significant at the peak of the solar cycle where three very large events occurred within six days. This clustering effect may be an important factor to consider in estimating maximum short-time doses.

4. The total annual flux appears to be in basic agreement with the variation in sunspot activity during the eleven year solar cycle. This implies that the cumulative flux

encountered during any particular mission will depend
on what portion of the solar cycle is encompassed.
Statistical prediction for cumulative flux should pre-
ferably be derived from a detailed time series analysis
of historical data on sunspots and solar flares, in
combination with extrapolation of solar activity levels
for the quieter years (1965-67) of the present cycle.

5. In order to ensure that the data on proton events for
the twentieth (current) solar cycle will satisfy minimum
requirements for valid statistical predictions, it is
essential that data gathering be controlled in accordance
with accepted principles of statistical design of experi-
ments. The recommendations for radiation monitoring made
by the Working Group on Radiation Problems of the National
Academy of Sciences/National Research Council do not
appear to be sufficiently comprehensive. The report states
that "Knowledge of the physical characteristics of the
radiation environments. . .is not immediately applicable
to radiation monitoring. Such information, therefore,
should be gathered and treated separately from the pro-
blem of crew safety." Statistically, it is believed
essential that the link between physical characteristics
and dosage be made as strong as possible. Moreover, the
link should be extended to include different types of
ground-based and satellite observations. More generally,
if the statistical point of view is to be adequately re-
flected in the current program on solar radiation, it is
recommended that experienced statisticians be directly
involved during the formative planning phases.

6. Development of appropriate statistical methods of
analysis--including goodness of fit, estimation, pre-
diction, and confidence limits--can be expected to pose
major problems in analysis. The varied solar phenomenon
should preferably be treated within an overall stochastic
process framework. It is likely that a fairly complex
statistical model will need to be employed in order to
reflect the variety of qualitative statistical character-
istics that have been pointed out in the previous analysis.

P. Gunther

P. Gunther

1033-PG-jr

Attachments
  References
  Appendix
  Tables 1 - 5
  Figures 1 - 8

BELLCOMM, INC.

REFERENCES

1. Radiobiological Factors in Manned Space Flight, W. H. Langham,
    National Academy of Sciences, National Research Council,
    Publication 1487, Washington, D. C., 1967.

2. R. H. Hilberg, "Solar Cosmic Ray Events," Bellcomm
    Memorandum for File, June 29, 1965.

3. W. R. Webber, "An Evaluation of the Radiation Hazard Due
    to Solar-Particle Events," Boeing Company, D2-90469,
    December 1963.

4. D. K. Bailey, "The Detection and Study of Solar Cosmic
    Rays by Radio Techniques," Journal of the Physical Society
    of Japan, vol. 17, Supp. A.I., Part I, January 1962.

5. J. L. Modisette, et al., "Model Solar Proton Environments
    for Manned Spacecraft Design," MSC, NASA TN D-2746,
    April 1963.

6. W. C. Lin and J. A. Van Allen, "Observation of Solar
    Cosmic Rays from October 13, 1959 to February 17, 1961
    with Explorer VII," State University of Iowa, SUI63-15.

7. M. Waldmeier, Results and Problems of Solar Research,
    2nd ed., Academic Publishing Company, Leipzig, Germany
    (in German), 1955.

8. A. Hald, Statistical Tables and Formulas, Wiley, 1962.

9. E. J. Gumbel, Statistics of Extremes, Columbia University
    Press, 1958.

10. E. J. Gumbel, "Statistical Theory of Extreme Values and
    Some Practical Applications," National Bureau of Standards,
    Applied Mathematics Series 33, February 12, 1954.

11. H. Jeffreys, Theory of Probability, 3d, ed., Oxford, Clarendon
    Press, 1961.

12. P. Gunther, "Essay on Bayesian Statistical Inference,"
    System Development Corporation Report, SP-1012/000/00,
    October 29, 1962.

BELLCOMM, INC.   - 2 -

13. P. Swerling, "Topics in Generalized Least Squares Estimation," *Journal of SIAM Applied Mathematics*, vol. 14, No. 5, September 1966.

14. D. B. Owen, *Handbook of Statistical Tables*, Addison-Wesley, 1962.

15. A. N. de Gaston, "Variation of Interplanetary Solar Cosmic Ray Radiation Hazard with Solar Cycle," Bellcomm Memorandum for File, October 25, 1966.

16. D. E. Guss, "Distribution in Heliographic Longitude of Flares Which Produce Energetic Solar Particles," *Physical Review Letters*, vol. 13, No. 12, p. 363, September 21, 1964.

BELLCOMM, INC.

APPENDIX

STATISTICAL CONFIDENCE DISTRIBUTIONS*

A.1 Basic Approach

Let $\underline{x} = (x_1, \cdots, x_n)$ be an observed sample (not necessarily independently or identically distributed) with probability density $f(\underline{x}/\theta)$ depending upon an unknown parameter θ. Suppose further that θ has an a priori distribution $f(\theta)$. The confidence density distribution $f(\theta/\underline{x})$ is then defined by Bayes formula

$$f(\theta/\underline{x}) = \frac{f(\underline{x}/\theta)f(\theta)}{\int_{-\infty}^{+\infty} f(\underline{x}/\theta)f(\theta)d\theta} \qquad (1)$$

In the important special case where θ is a location parameter for $\underline{x}$, if one assumes $f(\theta) = K$ (an arbitrary constant which cancels from numerator and denominator in (1)), then for a wide class of distributions, the inferences obtained from (1) are identical with those using classical methods. A discussion of the relation to the classical approach is given in reference 12. (See also reference 13.)

If θ is a scale parameter for $\underline{x}$, since log θ is then a location parameter for log $\underline{x}$, it follows that the appropriate a priori distribution for θ is $f(\theta) = K/\theta$.

If $\underline{x}$ is a scalar x (this includes the case where a scalar sufficient statistic** for θ exists), then if θ is a location parameter for x, one has

$$f(\theta/x) = f(x/\theta) \qquad (2)$$

---

*The origin of the ideas employed here is generally credited to H. Jeffreys.

**A sufficient statistic for a given family of distributions with unknown parameters is a function which contracts the observations without loss of information regarding the parameters.

Here $f(x/\theta)$ is considered as an ordinary function of the two variables x and θ. For the cumulative distribution one gets

$$F(\theta/x) = 1 - F(x/\theta) \qquad (3)$$

When θ is a scale parameter, (3) remains unchanged, while (2) requires a trivial modification.

A.2   <u>Uniform Distribution</u>

Let $\underline{x} = (x_1, \cdots, x_n)$ be a <u>random</u> sample from a uniform population ranging from 0 to θ. In this case θ is simultaneously a scale and a cutoff parameter. x_{max} is a sufficient statistic* for θ. Since each x_i is uniformly distributed, it follows that the distribution of x_{max} is given by

$$F(x_{max}/\theta) \equiv \text{Prob}(x_1 \leq x_{max}, \cdots, x_n \leq x_{max} \text{ given the cutoff } \theta)$$

$$= \left(\frac{x_{max}}{\theta}\right)^n , \quad 0 \leq x_{max} \leq \theta \qquad (3a)$$

The density is

$$f(x_{max}/\theta) = \frac{n \, x_{max}^{n-1}}{\theta^n} , \quad 0 \leq x_{max} \leq \theta \qquad (4)$$

---

*A sufficient statistic for a given family of distributions with unknown parameters is a function which contracts the observations without loss of information regarding the parameters.

Using $f(\theta) = K/\theta$ $(\theta > 0)$, since

$$\int_0^{\infty} f(x_{max}/\theta) \, f(\theta) d\theta = K/x_{max} \tag{5}$$

the confidence density distribution for θ is then, by (1),

$$f(\theta/x_{max}) = \frac{n \, x_{max}^n}{\theta^{n+1}} \, , \quad \theta \geq x_{max} \tag{6}$$

and the cumulative confidence distribution $F(\theta/x) \equiv \int_{-x}^{\theta} f(\theta/x) d\theta$
is

$$F(\theta/x_{max}) = 1 - \left(\frac{x_{max}}{\theta}\right)^n \quad \theta \geq x_{max} \tag{7}$$

in agreement with (3). This implies that the confidence limit
for specified P is given by

$$\theta_P = \left(1 - P\right)^{-\frac{1}{n}} \cdot x_{max} \tag{8}$$

Since this equation relates to the cutoff value, the appropriate
statistical intepretation is that the confidence is P that 100%
of the observations will be less than θ_P.

It is convenient to define Δ_P as the percentage in-
crease in x_{max} required to provide a confidence of P, i.e.,

$$\Delta_P = \frac{\theta_P - x_{max}}{x_{max}} \tag{9}$$

Then (8) can be written as

$$\Delta_P = \exp\left(\frac{1}{n} \ln \frac{1}{1-P}\right) - 1 \qquad (10)$$

If $n \gg \ln(1-P)^{-1}$, the first term in the expansion of (10) gives

$$\Delta_P \simeq \frac{1}{n} \ln \frac{1}{1-P} \qquad (11)$$

so that Δ_P is approximately inversely proportional to n. For
n = 24 and P = .99, the approximate value from (11) is .192
compared with the exact value of .212; for n = 24 and P = .90,
the approximate and exact values are .101 and .104, respective-
ly. Figure 8 plots Δ_P in equation (10) vs. n for P = .90, .95,
and .99.

The above results can be immediately generalized to
an arbitrary cumulative distribution $F(x/\theta)$ which is known
except for the (maximum finite) cutoff value θ. The probabil-
ity transformation (applied to both x and θ) can be used to
transform x to a uniform distribution. It follows that x_{max}
is a sufficient statistic for θ and the confidence distribution
is

$$F(\theta/x_{max}) = 1 - F^n(x_{max}/\theta) \ , \ \theta \geq x_{max} \qquad (12)$$

A.3 Truncated Distributions

The subscript T will denote truncation, with x_T being
the (known, left) truncation point. Letting $f_T(x)$ and $F_T(x)$
represent the truncated p d f and c d f of x, then

$$f_T(x) = \frac{f(x)}{1 - F(x_T)} \qquad (13)$$

$$F_T(x) = \frac{F(x) - F(x_T)}{1 - F(x_T)} \tag{14}$$

In the special case that x is uniformly distributed with cutoff θ and truncation point $x_T < \theta$, one gets

$$f_T(x/\theta) = \frac{1/\theta}{1 - x_T/\theta} = \frac{1}{\theta - x_T} , \quad x_T \leq x \leq \theta \tag{15}$$

$$F_T(x/\theta) = \frac{x/\theta - x_T/\theta}{1 - x_T/\theta} = \frac{x - x_T}{\theta - x_T} \tag{16}$$

Hence, x is uniformly distributed between x_T and θ, which fact is obvious intuitively. Equivalently, the reduced values $y_i = x_i - x_T$ are uniformly distributed between 0 and $\phi = \theta - x_T$. Also $y_{max} = x_{max} - x_T$ is a sufficient statistic for ϕ, and x_{max} is sufficient for θ. The density of x_{max} is given by

$$f_T(x_{max}/\theta) = \frac{n(x_{max} - x_T)^{n-1}}{(\theta - x_T)^n} , \quad x_T < x_{max} < \theta \tag{17}$$

(7) and (8) are similarly modified, with the confidence formula becoming

$$\theta_P = x_T + (1 - P)^{-1/n} (x_{max} - x_T) \tag{18}$$

Figure 8 is applicable to the present case with Δ_P in (9) now defined as

$$\Delta_P = (\theta_P - x_{max})/(x_{max} - x_T) \tag{19}$$

As a numerical example, equation (18) is applied to the flux data under the assumption that log flux (i.e., log N=x) is uniformly distributed with a finite but unknown cutoff θ. The confidence limits for cutoff flux are shown in Table 5 for the four cases treated in section 3.4, and for several choices of truncation point N_T, number n of observations, and confidence P. Regarding the choice of n, a difficulty arises because the observed fluxes within a cluster are not independent. (Equation (3a) used in deriving (8), requires independence.) Three different criteria have been used: (i) each event in a cluster is counted as a single observation, (ii) each event in a cluster is counted as 1/2 observation, and (iii) each cluster is counted as one observation.

Regarding the choice of truncation point, the uniform distribution invariably provided a better fit when the small and medium size events were not included. Hence, in addition to the truncation points used in fitting the distributions in Figures 4-7, a second value is used in Table 5 corresponding to the point where a sharp drop in the empirical density occurs. Finally, two confidence probabilities, 90% and 95%, are considered. The P% confidence estimate of cutoff flux, N_P, is simply $10^{\theta}P$, where θ_P is given by (18).

A.4 Confidence Distribution for the Pth Percentile of a Normal Distribution

The Pth percentile ξ_P of a normal distribution can be written as

$$\xi_P = \mu + u_P \sigma \qquad (20)$$

where u_P is the Pth percentile of the unit (normalized) normal distribution, e.g., $u_P = 2.326$ for P = 99%. If the confidence distribution for μ and σ are given, say from sample observations, it is desired to obtain the confidence distribution of ξ_P.

We suppose that the sample mean $\bar{x}$ is normally distributed about μ with variance σ^2/n, and that s^2/σ^2 has the χ^2 distribution with ν degrees of freedom, independent of the distribution of $\bar{x}$. (Except in special cases, ν is ordinarily equal to n-1.) The joint confidence distribution of μ and σ is the usual product of the normal and χ^2 distributions. Making

the substitution (20), one gets

$$f(\xi_P, \sigma/\overline{x}, s) = \left(\frac{n}{2\pi}\right)^{1/2} \sigma^{-1} e^{-\frac{n}{2\sigma^2}(\xi_P - u_P \sigma - \overline{x})^2}$$

$$\cdot \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} (S^2/2\sigma^2)^{\nu/2-1} s^2/\sigma^3 e^{-S^2/2\sigma^2} \qquad (21)$$

This can be simplified by letting

$$t = \frac{\sqrt{n\nu} \ (\xi_P - \overline{x})}{s} \qquad (22)$$

$$z = s/\sigma \ , \ 0 \le z < \infty \qquad (23)$$

Equation (21) then becomes

$$f(t,z) = \frac{(2\pi)^{1/2}}{\Gamma(\nu/2)2^{\nu/2-1}} \cdot (2\pi)^{-\frac{1}{2}} 2\nu^{-\frac{1}{2}} e^{-\frac{1}{2}\left(tz\nu^{-\frac{1}{2}} - u_P n^{\frac{1}{2}}\right)^2}$$

$$\cdot (2\pi)^{-\frac{1}{2}} e^{-z^2/2} z^{\nu-1} dt dz \qquad (24)$$

The density of t is obtained by integrating out z, i.e.,

$$f(t) = \int_0^\infty f(t, z) dz \qquad (25)$$

and the cumulative distribution is

$$F(t) = \int_{-\infty}^t \int_0^\infty f(t, z) dz dt \qquad (26)$$

We write the normal pdf and cdf as

$$\phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} \tag{27}$$

$$\Phi(x) = \int_{-\infty}^{x} \phi(y)dy \tag{28}$$

Note that

$$\int_{-\infty}^{x} a \phi(ay + b)dy = \int_{-\infty}^{ax+b} \phi(x) = \Phi(ax+b) \tag{29}$$

Substituting (24) and (29) into (26) and interchanging the order of integration gives

$$f(t) = (2\pi)^{-\frac{1}{2}} \Gamma(\nu/2)2^{\nu/2-1-1} \int_{0}^{\infty} \Phi(tz\nu^{-\frac{1}{2}} - u_p n^{\frac{1}{2}})\phi(z)z^{\nu-1}dx \tag{30}$$

Using (22), one can then easily obtain the confidence distribution of

$$\xi_p = \bar{x} + \frac{s}{\sqrt{n\nu}} t \tag{31}$$

Equation (30) is recognized as the non-central t distribution with ν degrees pf freedom and non-centrality parameter $u_p\sqrt{n}$. (See reference (14) for the formula and a table of numerical values.) Let t_γ be the value of t corresponding to confidence γ, i.e.,

$$F(t_\gamma) = \gamma$$

Note that t_γ depends upon ν and $u_P\sqrt{n}$. Then the γ confidence estimate of ξ_P, say $\xi_{P,\gamma}$ is, from (31),

$$\xi_{P,\gamma} = \bar{x} + \frac{t_\gamma}{\sqrt{n\nu}} \, s = \bar{x} + k_{P,\gamma} \, s \tag{32}$$

where

$$k_{P,\gamma} = t_\gamma (n\nu)^{-\frac{1}{2}}$$

The above result is identical with the classical solution.[14] The derivation given can be extended without difficulty to the case of a truncated normal distribution. However, the resulting distribution is not tabulated and would have to be evaluated numerically.

A.5   Type III Extreme Value Distribution

It is assumed that the random variable x has c d f and p d f given by

$$F(z) = e^{-\alpha(\theta-x)^k} \tag{33}$$

$$f(x) = k\alpha(\theta-x)^{k-1} e^{-\alpha(\theta-x)^k}, \quad -\infty < x < \theta \tag{34}$$

We consider first the special case where k is known to be unity, i.e.,

$$f(x) = \alpha e^{-\alpha(\theta-x)}, \quad -\infty < x < \theta \tag{35}$$

Although the density is exponential increasing up to the cutoff θ, and hence would not be used for solar events, the mathematical solution is instructive since it turns out to be completely analogous to the corresponding problem for the normal distribution.

The joint distribution of the random sample $\underline{x} = (x_1, \cdots, x_n)$ is

$$f(\underline{x}) = \alpha^n e^{-\alpha n(\theta - \overline{x})} \ , \ -\infty < x_n < \theta \qquad (36)$$

where for brevity we write x_n for $x_{max} = \max x_i$, and $\overline{x}$ is the sample mean $n^{-1} \Sigma x_i$. Note that θ and α are location and scale parameters, respectively, so that it is natural, as in the normal distribution, to use the a priori distributions $f(\theta) =$ constant, $f(\alpha) = $ constant. α^{-1}. Using the basic formula (1), the joint confidence distribution is

$$f(\theta, \alpha/\underline{x}) = \frac{\alpha^{n-1} e^{-\alpha n(\theta - \overline{x})}}{\displaystyle\int_{\theta = x_n}^{\infty} \int_{\alpha = 0}^{\infty} \alpha^{n-1} e^{-\alpha n(\theta - \overline{x})} \ d\alpha d\theta} \qquad (37)$$

With respect to α the integrand is recognized as the gamma function so that the denominator becomes, (letting C represent an appropriate constant depending on n)

$$\iint \cdots \ d\alpha d\theta = \int_{x_n}^{\infty} \frac{C}{(\theta - \overline{x})^n} \ d\theta$$

$$= \int_{x_n}^{\infty} \frac{C(x_n - \overline{x})^{-n}}{\left|1 + \dfrac{\theta - x_n}{x_n - \overline{x}}\right|^n} \ d\theta \simeq \frac{C}{(x_n - \overline{x})^{n-1}} \qquad (38)$$

Hence

$$f(\theta, \alpha/\underline{x}) = C(x_n - \overline{x})^{n-1} \ \alpha^{n-1} \ e^{-\alpha n(\theta - \overline{x})} \qquad (39)$$

The confidence distribution of θ is obtained by integrating out α in (39), i.e.,

$$f(\theta/\underline{x}) = \int_0^{\infty} f(\theta,\alpha/\underline{x})d\alpha = \frac{C(x_n-\bar{x})^{n-1}}{(\theta-\bar{x})^n} = \frac{C}{(x_n-\bar{x})\left(1+\dfrac{\theta-x_n}{x_n-\bar{x}}\right)^n} \qquad (40)$$

This density is recognized as a known tabulated distribution closely related to Student's t-distribution used in connection with the normal distribution. Specifically, the quantity

$$(n-1)\left(\frac{\theta-x_n}{x_n-\bar{x}}\right)$$

has the F-distribution with 2 and $2n - 2$ degrees of freedom.

One can similarly derive the density for α alone, i.e., $f(\alpha/\underline{x})$. In fact, $\alpha n(x_n-\bar{x})$ has the gamma distribution with n-1 degrees of freedom.

One can also derive the above results, which are undoubtedly well-known, by the usual classical methods. One shows first that x_n and $x_n-\bar{x})$ are independently distributed,[*] that $2n\alpha(\theta-x_n)$ has a χ^2 distribution with 2 degrees of freedom, and that $2\alpha n(x_n-\bar{x})$ has a χ^2 distribution with 2n-2 degrees of freedom. It follows then that the ratio of the two quantities, weighted inversely to their degrees of freedom, namely

---

[*]One can also see from (36) that these are joint sufficient statistics for θ and α.

$$\frac{\frac{1}{2} \cdot 2n\alpha(\theta-x_n)}{\frac{1}{2n-2} \cdot 2n\alpha(x_n-\bar{x})} = \frac{(n-1)(\theta-x_n)}{(x_n-\bar{x})} \quad ,$$

has the $F_{2,2n-2}$ distribution.

     If the same type of argument is employed for general k, i.e., using equation (34), one gets (where C now depends on k as well as n)

$$\int_0^\infty f(\underline{x}/\theta,\alpha,k)d\alpha = \frac{C\pi(\theta-x_i)^{k-1}}{\left[\Sigma(\theta-x_i)^k\right]^n}$$

Letting $x_i' = x_n-x_i$ (i = 1, $\cdots$, n) and $y = \theta - x_n$, one can write

$$\int_{x_n}^\infty \frac{\pi(\theta-x_i)^{k-1}}{\left[\Sigma(\theta-x_i)^k\right]^n} d\theta = \int_0^\infty \frac{\pi(y+x_i')^{k-1}}{\left[\Sigma(y+x_i')^k\right]^n} dy = g(x_i', \cdots, x_{n-1}) \text{ say}$$

observe that when k = 1, $g = (\Sigma x_i')^n$. The confidence distribution for θ is thus

$$F(\theta/\underline{x},k) = \frac{\int_{x_n}^\theta \frac{\pi(y+x_i')^{k-1}}{\left[\Sigma(\theta-x_i)^k\right]^n} d\theta}{\int_{x_n}^\infty \frac{\pi(\theta-x_i)^{k-1}}{\left[\Sigma(\theta-x_i)^k\right]^n} d\theta}$$

$$= g(x_i', \cdots, x_{n-1}')^{-1} \int_0^{\theta-x_n} \frac{\pi(y+x_i')^{k-1}}{\left[\Sigma(y+x_i')^k\right]^n} dy$$

Given the x_i, this expression can be numerically integrated to obtain the confidence value for θ corresponding to any specified value P. It is believed that the result is statistically optimum. The derivation is easily modified to apply to the Type I and Type II extreme value distributions.

Ordinarily, the parameter k will not be known but must be inferred from the data as was done for θ and α. Unfortunately, k cannot be transformed into a location parameter, so that there does not seem to be a "natural" a priori distribution for k. A similar situation arises with the gamma distribution when the number of degrees of freedom is unknown. In this instance, reasonable inferences have been derived assuming that the a priori distribution of the parameter is uniform over the positive integers. For the present problem, since k need not be an integer, a reasonable a priori distribution would appear to be $f(k)$ = constant for $k \geq 1$. To get then the confidence distribution for θ, an additional integration would be required using

$$f(\theta/\underline{x}) = \int f(\theta/\underline{x},k) \; f(k) \, dk$$

$$= \lim_{A \to \infty} \frac{1}{A-1} \int_1^A f(\theta/\underline{x},k) \, dk$$

TABLE 1. SOLAR PROTON EVENTS FROM 1956 TO 1961[†]

Year	Date	Integrated Flux (Protons/cm^2) $N(>30\ Mev)$	$N(>100\ Mev)$
1956	Feb. 23	1.0×10^9	3.5×10^8
	Mar. 10*	1.1×10^8 (*)	
	Aug. 31	2.5×10^7	6.0×10^6
	Nov. 13*	1.1×10^8 (*)	
1957	Jan. 20	2.0×10^8	7.0×10^6
	Apr. 3* }	5.6×10^7 (*)	
	Apr. 6* }	3.8×10^7 (*)	
	June 22*	1.7×10^8 (*)	
	July 3	2.0×10^7	-----
	Aug. 9	1.5×10^6	-----
	Aug. 29 }	1.2×10^8	3.0×10^6
	Aug. 31* }	5.3×10^7 (*)	
	Sept. 2* }	1.4×10^7 (*)	
	Sept.21	1.5×10^6	-----
	Oct. 20	5.0×10^7	1.0×10^7
	Nov. 4	9.0×10^6	-----
1958	Feb. 9	1.0×10^7	-----
	Mar. 23 }	2.5×10^8	1.0×10^7
	Mar. 25* }	7.8×10^8 (*)	
	Apr. 10	5.0×10^6	-----
	July 7	2.5×10^8	9.0×10^6
	Aug. 16 }	4.0×10^7	1.6×10^6
	Aug. 22 }	7.0×10^7	1.8×10^6
	Aug. 26 }	1.1×10^8	2.0×10^6
	Sept.22	6.0×10^6	1.0×10^5

Year	Date	Integrated Flux (Protons/cm^2) $N(>30\ Mev)$	$N(>100\ Mev)$
1959	Feb. 13*	2.8×10^7 (*)	
	May 10	9.6×10^8	8.5×10^7
	June 13	8.5×10^7	-----
	July 10 }	1.0×10^9	1.4×10^8
	July 14 }+	1.3×10^9	1.0×10^8
	July 16 }+	9.1×10^8	1.3×10^8
	Aug. 18	1.8×10^6	-----
1960	Jan. 11	4.0×10^5	-----
	Mar. 29*	2.7×10^7 (*)	
	Apr. 1 }	5.0×10^6	8.5×10^5
	Apr. 5 }+	1.1×10^6	-----
	Apr. 28 }	5.0×10^6	7.0×10^5
	Apr. 29 }-	7.0×10^6	-----
	May 4 }+#	6.0×10^6	1.2×10^6
	May 6 }-	4.0×10^6	-----
	May 13 }-	4.0×10^6	4.5×10^5
	June 1	4.0×10^5	-----
	Aug. 12	6.0×10^5	-----
	Sept. 3	3.5×10^7	7.0×10^6
	Sept.26	2.0×10^6	1.2×10^5
	Nov. 12 }	1.3×10^9	2.5×10^8
	Nov. 15 }+	7.2×10^8	1.2×10^8
	Nov. 20 }+	4.5×10^7	8.0×10^6
1961	July 11 }	3.0×10^6	2.4×10^5
	July 12 }+	4.0×10^7	1.0×10^6
	July 18 }+	3.0×10^8	4.0×10^7
	July 20 }+	5.0×10^6	9.0×10^5
	Sept.10*	3.7×10^7 (*)	
	Sept.28	6.0×10^6	1.1×10^6

†From reference 5. + between two events signifies that both originated from approximately the same solar region; - signifies different regions.

*Bailey event.

#The May 4, 1960 event is correlated with the April 28, 1960 event.

TABLE 2.   SUMMARY OF ANNUAL NUMBER AND FLUX OF SOLAR PROTON EVENTS

| Year | Bailey Events | Webber Events (Log Flux > 30 Mev) | | | | | | | | Total |
	5.5-5.9	6.0-6.4	6.5-6.9	7.0-7.4	7.5-7.9	8.0-8.4	8.5-8.9	9.0-9.5		
1956	2				1				1	4
1957	5		2	1	1	1	2			12
1958	1			2	1	2	3			9
1959	1		1			1		2	2	7
1960	1	3	2	6		2		1	1	16
1961	1		1	2		1	1			6
TOTAL	11	3	6	11	3	7	6	3	4	54

TABLE 3.   TOTAL QUARTERLY FLUX

Quarter		No. of Bailey Events	Total No. of Events	No. of Clusters	$N(>30 \text{ Mev})$[†]	$N(>100 \text{ Mev})$
	I	1	2	2	1×10^9	3.5×10^8
1956	II	0	0	0	0	0
	III	0	1	1	2.5×10^7	6×10^6
	IV	1	1	1	1.1×10^6	(**)
	I	0	1	1	2×10^8	7×10^5
1957	II	3	3	2	2.64×10^6	(**)
	III	2	6	4	2×10^8	3×10^6
	IV	2	2	2	5.9×10^7	1×10^7
	I	1	3	2	2.7×10^8	1×10^7
1958	II	0	1	1	5×10^6	--
	III	0	5	4	4.76×10^8	1.44×10^7
	IV	0	0	0	0	0
	I	1	1	1	2.8×10^5	(**)
1959	II	0	2	2	1.045×10^9	8.5×10^7(+)
	III	0	4	2	3.21×10^9	3.7×10^8
	IV	0	0	0	0	0
	I	1	2	1*	6.7×10^5	--
1960	II	0	8	3	3.24×10^7	2.75×10^6
	III	0	3	3	3.76×10^7	7.12×10^6
	IV	0	3	1	2.065×10^9	3.78×10^8
	I	0	0	0	0	0
1961	II	0	0	0	0	0
	III	1	6	3	3.544×10^8	4.22×10^7

†Bailey 30 Mev flux reduced by factor of 100.
*The March 29 event is considered part of the April 1 and April 5 cluster.
**All Bailey events--no 100 Mev flux available.

TABLE 4. X^2 VALUES AND PROBABILITIES FOR GOODNESS OF FIT

Density	Ind'l Events-- 30 Mev (Fig.4)			Clusters-- 30 Mev (Fig.5)			Ind'l Events-- 100 Mev (Fig.6)			Clusters-- 100 Mev (Fig.7)		
	X^2	Deg. Freedom	Prob.	X^2	Deg. Freedom	Prob.	X^2	Deg. Freedom	Prob.	X^2	Deg. Freedom	Prob.
UNIFORM	3.98	5	.55	6.77	4	.16	15.45	4	.005	9.34	3	.025
LINEAR: Last Interval 0	8.14	6	.22	2.44	5	.78	7.59	5	.19	6.27	4	.19
LINEAR: Last 2 Int'ls Combined	.37	5	.996	1.60	4	.80	6.75	4	.16	4.73	3	.19
NORMAL: Last Interval 0	12.18	5	.04	2.82	4	.59						
NORMAL: Last 2 Intervals Combined	2.14	4	.71	1.60	3	.65						

TABLE 5. CONFIDENCE ESTIMATES FOR CUTOFF FLUX (UNIFORM DISTRIBUTION)

Flux Data	N_{max}	Truncation	N_T	n Criterion for clustered events	n	Cutoff Flux Prob=.90 $N_{.90}$	Prob=.95 $N_{.95}$
30 Mev Individual Events	1.3×10^9	All events	2.5×10^5	Each event=1	54	1.9×10^9	2.1×10^9
				Each event=1/2	42	2.1×10^9	2.4×10^9
				Cluster =1	39	2.2×10^9	2.6×10^9
		Small events excluded	1.4×10^7	Each event=1	22	2.1×10^9	2.5×10^9
				Each event=1/2	17	2.5×10^9	3.0×10^9
				Cluster =1	16	2.6×10^9	3.3×10^9
30 Mev Clusters	3.2×10^9	All events	2.5×10^5	Cluster =1	39	5.7×10^9	6.8×10^9
		Small events excluded	1.4×10^7	Cluster =1	16	7.0×10^9	8.2×10^9
100 Mev Individual Events	3.5×10^8	Small events excluded	4.0×10^5	Each event=1	34	5.6×10^8	6.5×10^8
				Each event=1/2	26.5	6.5×10^8	7.9×10^8
				Cluster =1	25	6.7×10^8	8.3×10^8
		Medium events excluded	1.3×10^6	Each event=1	21	6.8×10^8	8.5×10^8
				Each event=1/2	17	8.0×10^8	1.0×10^9
				Cluster =1	16	8.5×10^8	1.1×10^9
100 Mev Clusters	3.8×10^8	Small events excluded	4.0×10^5	Cluster =1	25	7.0×10^8	9.1×10^8
		Large events Only	1.6×10^7	Cluster =1	5	2.4×10^9	5.1×10^9
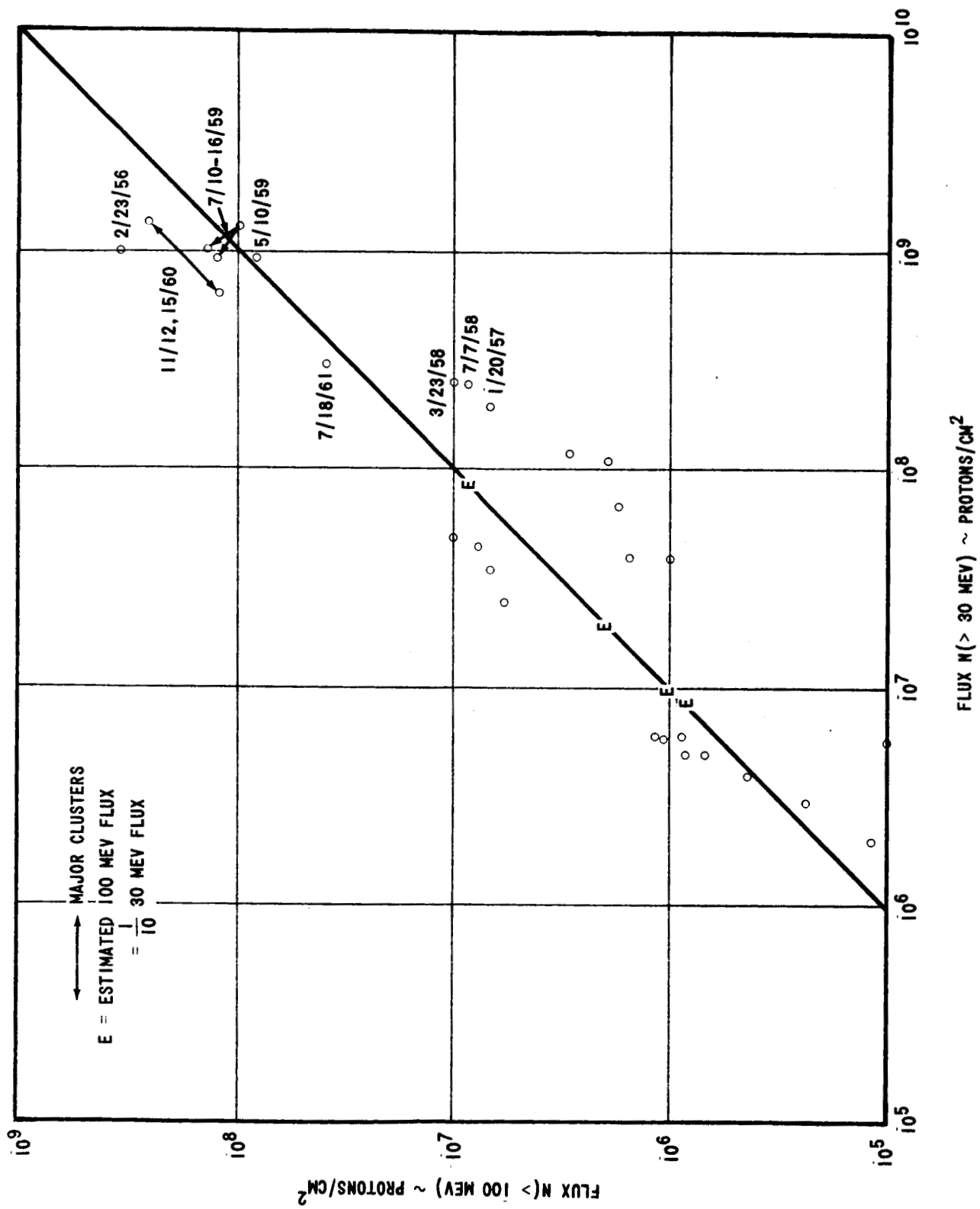
FIGURE I - SCATTER DIAGRAM OF FLUX FOR ENERGY > 100 MEV vs. FLUX FOR ENERGY > 30 MEV

SMOOTHED SUNSPOT NUMBER

NUMBER OF FLARES

SOLAR PROTON FLUX

NUMBER OF FLARES
(DeGASTON, REF. 15)

SMOOTHED SUNSPOT NO.
(DeGASTON, REF. 15)

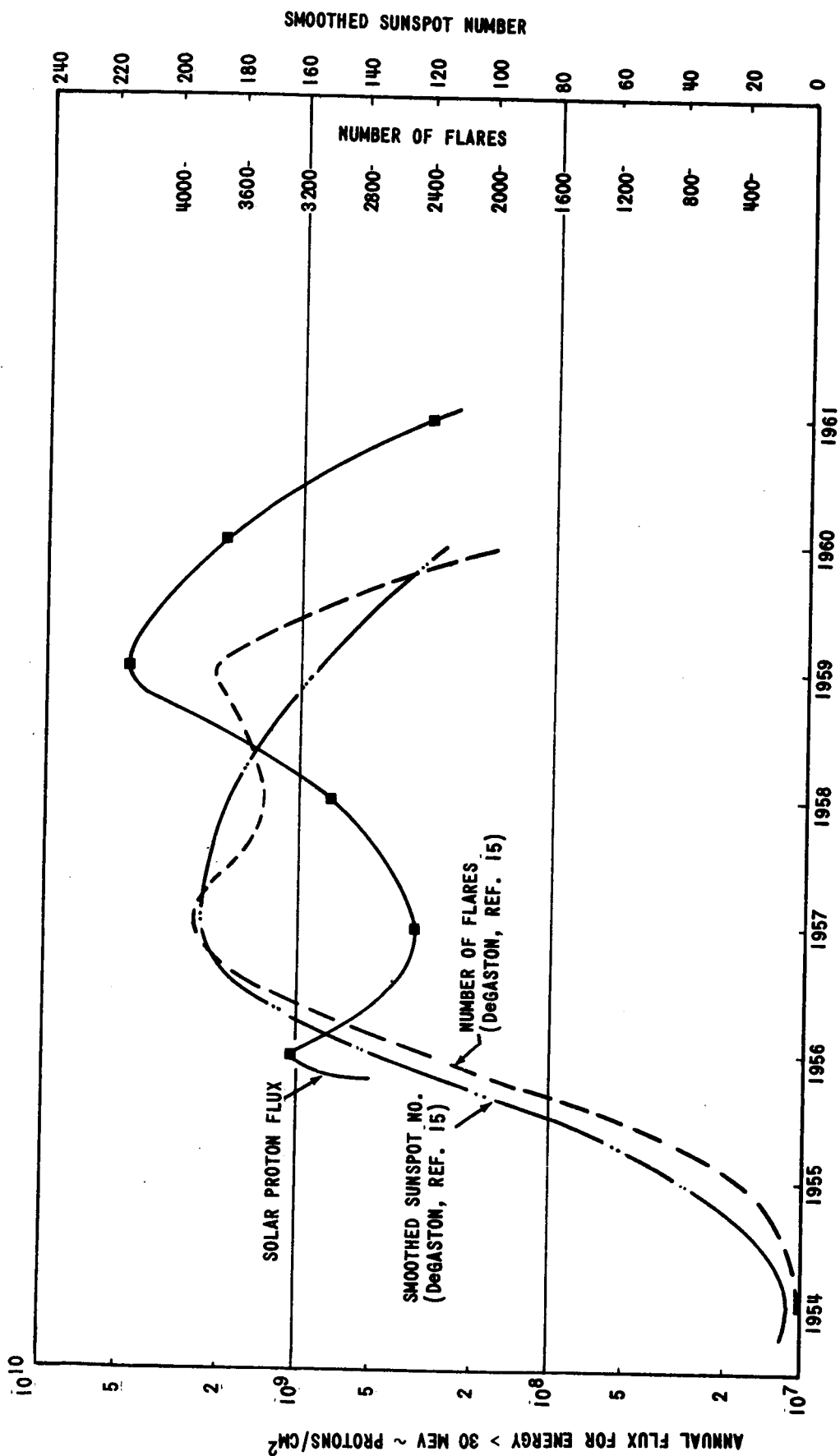ANNUAL FLUX FOR ENERGY > 30 MEV ~ PROTONS/CM^2

FIGURE 2 – COMPARISON OF PROTON EVENT FLUX, SUNSPOTS, AND SOLAR FLARES

SUNSPOT NUMBER CYCLE
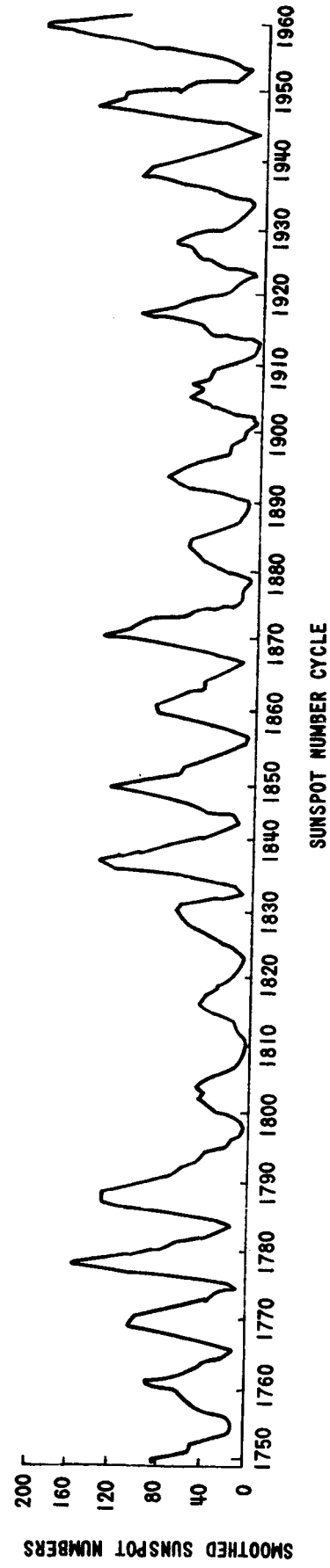
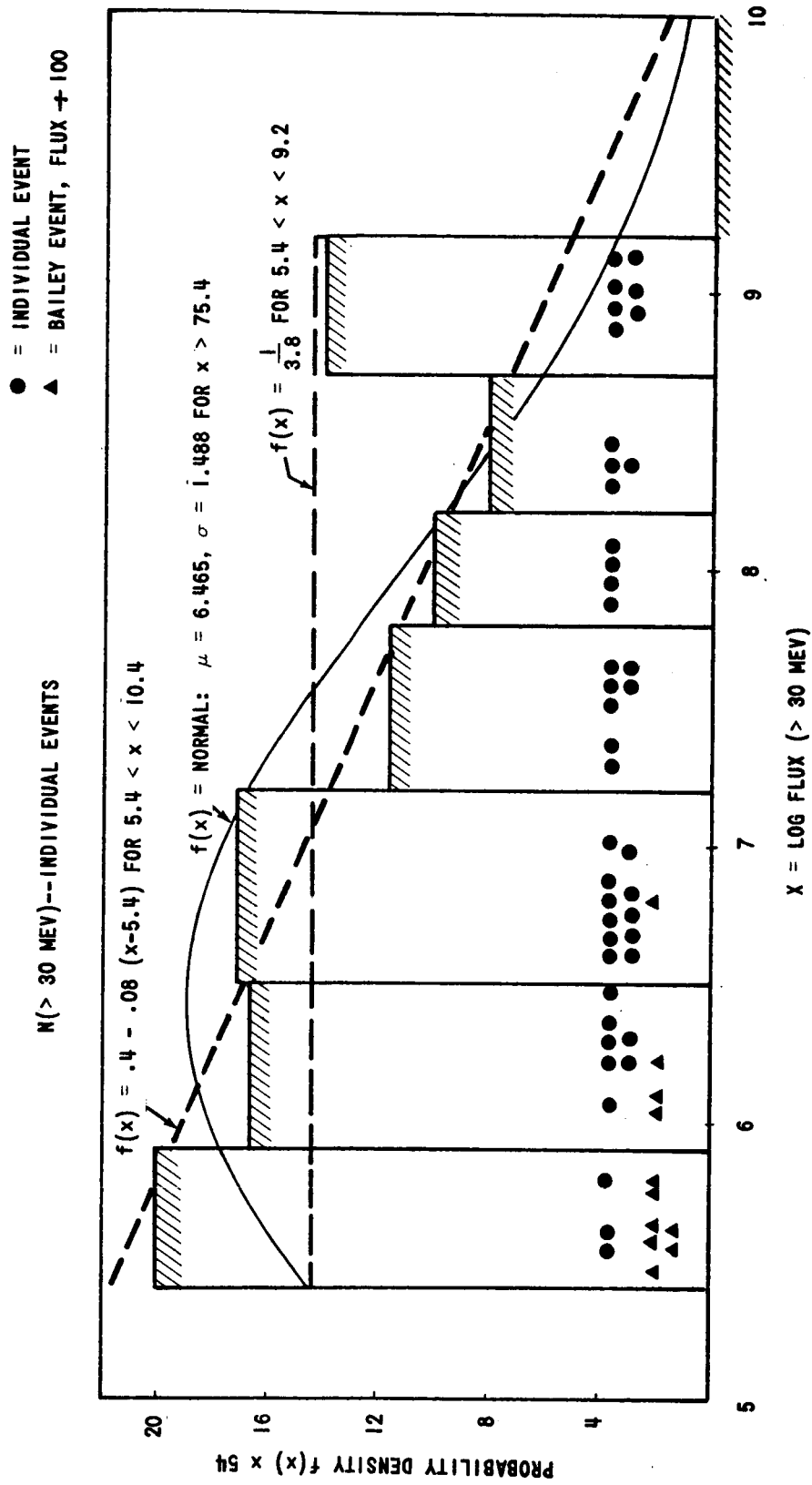FIGURE 3 - SMOOTHED SUNSPOT NUMBERS FROM 1750 TO 1960

FROM M. WALDMEIER, (REFERENCE 3)

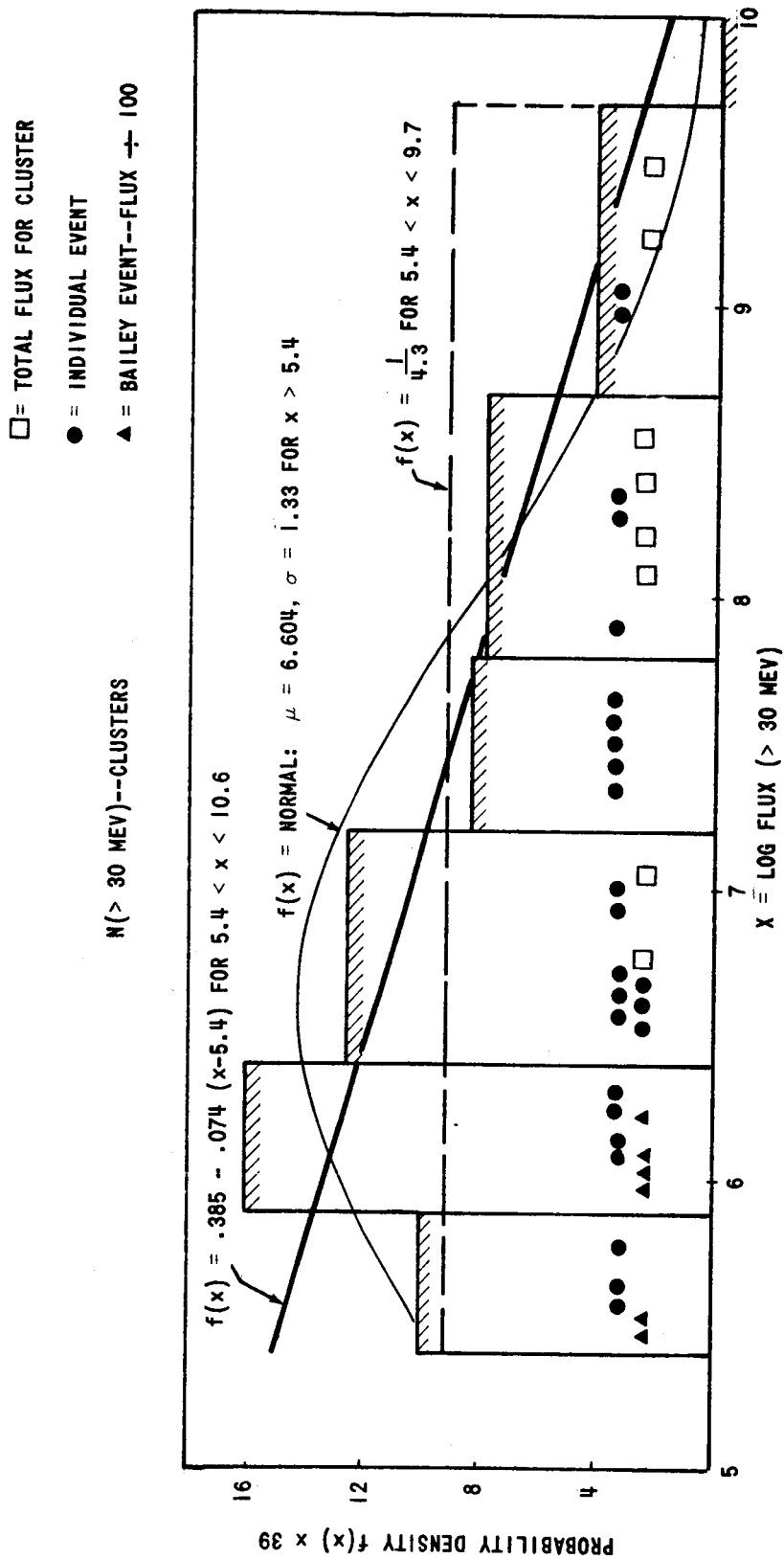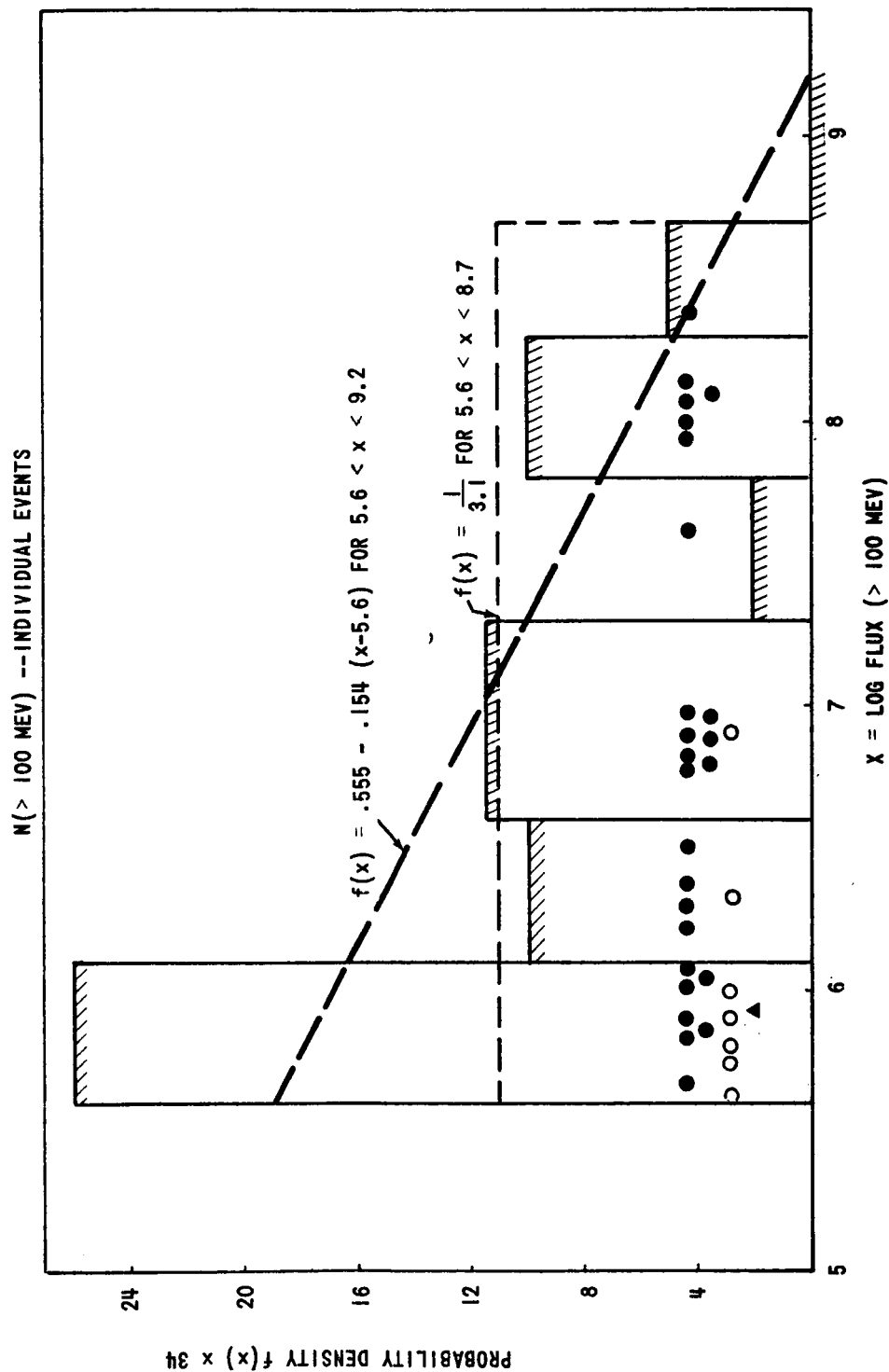FIGURE 4 -- EMPIRICAL AND FITTED PROBABILITY DENSITIES OF FLUX FOR INDIVIDUAL EVENTS--
ENERGY > 30 MEV

FIGURE 5 -- EMPIRICAL AND FITTED PROBABILITY DENSITIES OF FLUX (SUMMED FOR CLUSTERS)--ENERGY > 30 MEV

FIGURE 6 - EMPIRICAL AND FITTED PROBABILITY DENSITIES OF FLUX FOR INDIVIDUAL EVENTS - ENERGY > 100 MEV
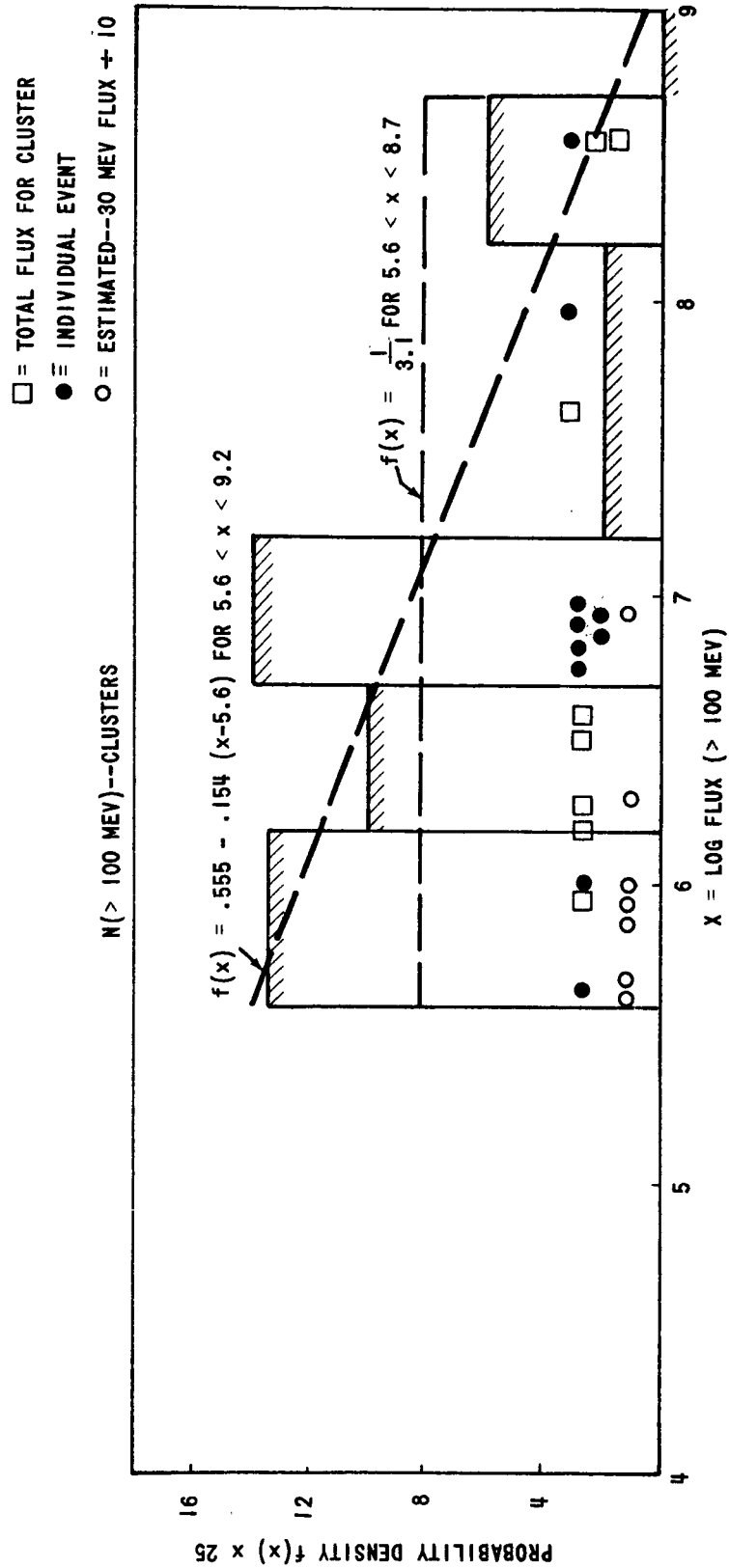
FIGURE 7 — EMPIRICAL AND FITTED PROBABILITY DENSITIES OF FLUX (SUMMED FOR CLUSTERS)—ENERGY > 100 MEV

$$f(x) = \frac{1}{\theta - x_T}, \quad x_T < x < \theta$$

$$\Delta_P = \frac{\theta_P - \hat{x}_{max}}{x_{max} - x_T} = (1-P)^{-\frac{1}{N}} - 1 \cong \frac{1}{N} \, LN \, \frac{1}{1-P}$$

P = 99%

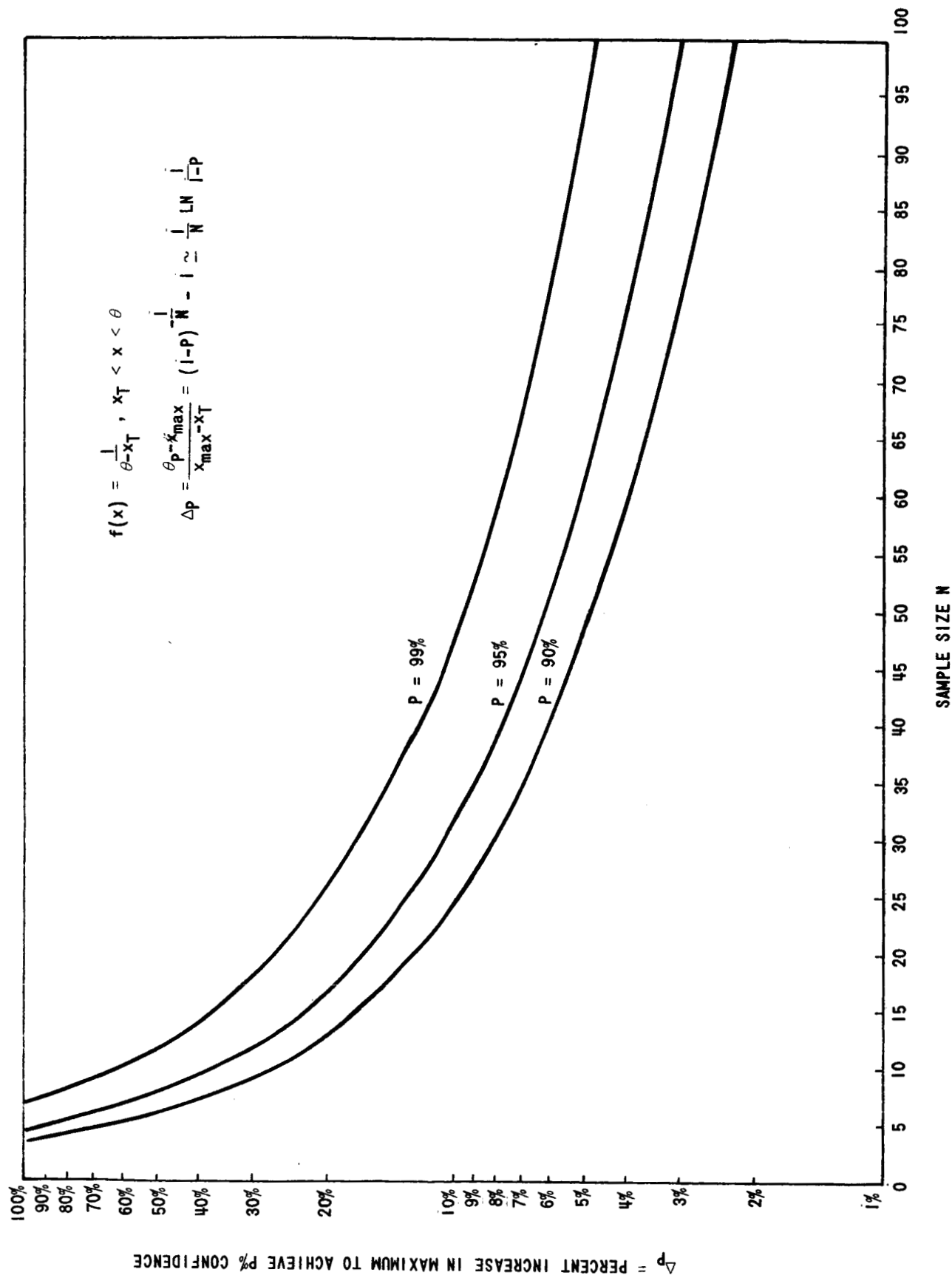P = 95%

P = 90%

SAMPLE SIZE N

FIGURE 8 - CONFIDENCE MARGIN BEYOND OBSERVED MAXIMUM FROM A UNIFORM POPULATION
VS. SAMPLE SIZE N AND CONFIDENCE P

Δ_P = PERCENT INCREASE IN MAXIMUM TO ACHIEVE P% CONFIDENCE